



**HAL**  
open science

## Transposable elements

Alexandre Hayward, Clement Gilbert

► **To cite this version:**

Alexandre Hayward, Clement Gilbert. Transposable elements. *Current Biology - CB*, 2023, 32 (17), pp.PR904-R909. 10.1016/j.cub.2022.07.044 . hal-03799727

**HAL Id: hal-03799727**

**<https://hal.science/hal-03799727v1>**

Submitted on 10 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Primer

# Transposable elements

Alexander Hayward<sup>1</sup> and Clément Gilbert<sup>2</sup>

<sup>1</sup>Centre for Ecology and Conservation, University of Exeter, Cornwall Campus Treliever Road Penryn, Cornwall, TR10 9FE, UK.

<sup>2</sup>Laboratoire Evolution, Génomes, Comportement, Ecologie, CNRS Université Paris-Saclay UMR 9191 IRD UMR 247, Avenue de la Terrasse, Bâtiment 13, Boite Postale 1, 91198 Gif sur Yvette, France.

Email: [alex.hayward@exeter.ac.uk](mailto:alex.hayward@exeter.ac.uk), [clement.gilbert@egce.cnrs-gif.fr](mailto:clement.gilbert@egce.cnrs-gif.fr).

Transposable elements are known by many names, including ‘transposons’, ‘interspersed repeats’, ‘selfish genetic elements’, ‘jumping genes’, and ‘parasitic DNA’, but here we will refer to them simply as TEs. Many biologists will have heard of TEs and their ability to transpose (change position) within the genome. But fewer may be aware of their varied influences on host biology, including contributions to the evolution of diverse host traits such as internal gestation, memory, colouration, and adaptive immunity. TEs are a near ubiquitous feature of eukaryotic genomes, and they often comprise a substantial proportion of total genomic content. Consequently, TE genes are considered among the most abundant coding sequences in nature. Recent advances in genome sequencing offer a golden age for TE research, providing opportunities to greatly improve understanding of interactions with host evolution and disease. However, significant challenges remain in our ability to detect and analyse TEs, which impair efforts to decipher their evolution, characterise their diversity, and elucidate their myriad host influences. Below, we summarise key aspects of TE biology and discuss major outstanding research questions.

## Transposable elements as genomic parasites

Active TEs play no inherent role in host biology, instead operating as genomic parasites. Just like other parasites, the fitness of TEs depends upon their ability to evade host defences, proliferate, and spread to other individuals. Likewise, in pursuing their own selfish evolutionary trajectories, TEs impart various negative impacts on their hosts, which range considerably in their costs. While the costs associated with inducing a host to synthesise an individual TE are nearly neutral, if transposition occurs into an essential host gene or regulatory region, insertional mutagenesis can be highly deleterious or even fatal to the host. Meanwhile, as TEs proliferate within a genome, the accumulation of repetitive sequences elevates the risk of ectopic recombination (recombination at non-homologous loci), increasing genomic instability and its accompanying negative effects. TEs are also directly associated with several human diseases, including muscular dystrophy, haemophilia, and several forms of cancer.

Given the costs outlined above, host genomes have evolved strategies to control TE activity. These include epigenetic mechanisms that alter chromatin conformation to reduce TE expression, and targeted degradation of TE transcripts. Consequently, after an initial period of proliferation, the activity of a given TE tends to decline or halt altogether. Strategies of TE repression may differ according to cell identity, such as between host

somatic and germ-line cells (the integrity of which is vital for the continuity of a host lineage). However, much remains unclear about the precise mechanisms used to control TE activity, and how these vary within and among hosts. Additionally, while TE activity can often be held in check relatively successfully, hosts have been remarkably unsuccessful at excluding TEs, as almost all sequenced eukaryotic genomes contain them, frequently in large proportions.

A major discovery of the postgenomic era was that genome size varies greatly among species, and that this is partially explained by substantial variation in TE content (e.g. maize 84%, pig 40%, *Drosophila* 20%, chicken 10%, brewer's yeast 3%). The reasons behind this are not well understood, and further research is needed to determine the various factors that influence TE accumulation and persistence. In humans, TEs account for approximately two thirds of our 3.2Gb genome, dwarfing the 1-2% encoding our genes. Furthermore, TEs may account for more of the genome than we currently recognise. TEs gradually accumulate mutations at the background rate of mutation in a genome. As a result, many TE sequences are 'genomic fossils' that can no longer mobilise. Therefore, over evolutionary timescales, such mutational processes can erode the genetic hallmarks diagnostic of TE sequences, until they are no longer recognisable. Depending on the rate of sequence turnover in a genome, heavily degraded TE sequences may thus comprise a significant proportion of the so-called 'dark matter of the genome'.

### **Transposable elements as catalysts of host evolvability**

Initially, TEs were often considered to be little more than junk DNA that littered host genomes. However, it has become clear that TEs have been recruited on many occasions, and for varied host functions, over the course of eukaryotic evolution. The recruitment of a TE for host purposes typically involves its inactivation, melding the persistence of its sequences to a fixed locus within the host genome, and ending its separate evolutionary trajectory as a genomic parasite (exceptions include telomeric TEs in *Drosophila* and those involved in antiviral immunity in mosquitos).

TEs can be harnessed for host functions in many ways. As well as simply knocking out a gene via insertional mutagenesis, TEs can donate coding sequence, modify gene splicing, mediate gene duplication, and alter chromatin conformation. Additionally, many TEs contain regulatory elements, and they are frequently implicated in host gene regulation. For example, it is estimated that around a quarter of human promoter regions contain TE-derived sequences. Further, given that copies of the same TE often disperse across the genome, TEs are well positioned to facilitate the re-wiring of whole host gene-regulatory networks. Prominent cases where TEs may have contributed in this way include regulatory networks governing embryonic stem cells, pregnancy, brain development, and innate immunity.

Recruitment of TE sequences for host purposes is frequently described as 'co-option'. In some cases, co-opted TE sequences undergo a period of adaptation, referred to as 'molecular domestication', during which they are fine-tuned to their new host function. In other cases, TEs are co-opted in a 'plug-and-play' manner, without need for further optimisation. The potential for co-option is augmented by the great diversity of molecular components that TEs have evolved since their ancient origins. It is further aided by their modular structure, whereby distinct and separable subunits code for independent tasks, enabling efficient host acquisition of beneficial functions. Consequently, TEs represent a cache of genetic innovation, which when re-purposed, can profoundly influence host evolution, facilitating adaptation and in some cases major evolutionary transitions (Figure 1).

Numerous examples exist where TEs have played significant roles during vertebrate evolution. These include contributions to functions of major importance in our lives, such as defending against disease, incubating young, and memory formation. The domestication of recombination activating genes from transposase genes was an important step in the evolution of adaptive immunity in jawed vertebrates, facilitating V(D)J recombination during B and T cell maturation. Syncytin genes, which are essential for placental development by facilitating cellular fusion in the layer separating maternal and fetal blood, have multiple independent evolutionary origins from diverse endogenous retrovirus envelope genes. *Arc* genes, which originate from LTR retrotransposon *group-specific antigen* genes, encode proteins that transport mRNA among synapses, performing key functions in memory formation and brain development. A TE insertion even appears to have provided the genetic mechanism by which apes lost their tails! Specifically, insertion of a SINE TE into an intron of the *brachyury* gene in the ancestor of hominoids is implicated in causing alternative splicing and tail-loss.

TE mobilisation rates have rarely been estimated and they probably vary greatly according to TE and host type. Generally, a particular TE copy will only transpose once per large number of host generations. However, certain conditions, such as large TE loads, large host population sizes, and strong selection, may increase the likelihood that novel TE insertions arise and are recruited for host purposes. Meanwhile, there is evidence that TE activity can increase considerably during periods of host stress, such as that arising from exposure to pathogens or xenobiotic agents. Coupled with strong selection for resistance, stress-induced increases in TE activity may potentially provide an engine of evolvability, whereby beneficial mutations arise and spread through a population. Therefore, testing the extent to which bursts of elevated TE activity contribute to host evolvability is relevant to diverse applied fields, including the design of pest management regimes to minimize resistance evolution, the conservation of endangered species, and predicting how species will cope with climate change.

### **Transposable element diversity**

There are two main divisions of TEs: retrotransposons, which use an RNA intermediate for transposition, and DNA transposons, which use a DNA intermediate (Figure 2).

Retrotransposons are divided further into: long terminal repeat (LTR) elements, DIRS-like elements (DLEs), Penelope-like elements (PLEs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs). While DNA transposons are divided into: DDE, Crypton, Helitron, and Maverick/Polinton elements (Figure 2).

Retrotransposons employ replicative transposition via a ‘copy-and-paste’ mechanism. DNA transposons, apart from Helitrons (which employ a replicative ‘peel-and-paste’ mechanism), transpose via a non-replicative ‘cut-and-paste’ mechanism. However, DNA transposons can proliferate via other means, such as recombinational repair or transposition across the replication fork. Eukaryotic genomes typically contain TEs belonging to several TE types, but their relative abundance can vary dramatically, even among closely related taxa.

Nevertheless, several broadscale patterns exist. For example, LTR retrotransposons are particularly abundant in plants, while LINEs and SINEs often dominate TE content in vertebrate genomes.

The lowest level of TE taxonomy is the family, typically defined by applying a threshold such as the “80-80” rule: TEs belong to the same family if their sequences share more than 80% sequence similarity over 80% of their length. TE families are usually represented as consensus sequences, which summarise variation among a set of TE copies extracted from a given genome. However, families assigned using the 80-80 rule do not

necessarily correspond to monophyletic groups, and classification approaches that include estimates of the phylogenetic relationships among TE sequences are more evolutionarily robust.

TEs lack an official nomenclature, and the names bestowed upon them by researchers vary from the utilitarian (DNA4-11\_CGi), to the fantastical (*Anakin*, *Helraiser*, *Idéfix*, *Quasimodo*, *Zaphod*), historical (*Athilla*, *HMSBeagle*, *Merlin*, *Samuri*) and mythological (*Athena*, *Icarus*, *Poseidon*, *Odin*). Names alluding to their mobility (*Mariner*, *Castaway*, *Stowaway*) and ability to ‘jump’ within the genome (*Kangaroo*, *Pogo*, *Tigger*) are also common. While less romantic, the adoption of a standardised reference-based naming convention may be beneficial in the future, especially since just a tiny fraction of eukaryotic diversity has been screened for TEs.

Researchers use a combination of approaches to detect TEs in a genome. Library-based methods perform sequence comparisons against databases of known elements. However, since many TEs remain unknown, library-based annotations typically require supplementation by ‘*de novo*’ approaches. These identify novel TEs using either: (i) sequence signatures associated with TE structure; or (ii) sequence repetitiveness (TEs are characteristically repetitive). In practice, most TE annotations are automated, relying on software pipelines employing combinations of approaches. However, many such pipelines exist, with different strengths and weaknesses, and performance can vary greatly. Currently, high-quality TE annotations, that exclude multicopy host genes and other repetitive non-TE sequences, but include full-length instead of partial TE models, are typically only achievable via manual curation. Yet given the time and effort required to curate TE models, most TE annotations are automated, and the quality control of reference sequences deposited into TE databases is an issue of growing concern.

### **The tangled origins of transposable elements**

TEs do not share a single evolutionary origin. Moreover, the distinction between viruses and TEs is blurred in several cases, complicating attempts to elucidate their respective histories. Viruses are often defined as infectious particles (virions) that replicate within the living cells of organisms. Meanwhile, TEs are considered mobile genetic elements that are not infectious and are restricted to individual host cells. However, biology frequently eludes convenient definitions, and the precise situation is not quite so neat. LTR retrotransposons and retroviruses almost certainly share a common evolutionary origin, but the former are considered TEs and the latter are considered viruses. Both groups possess a *group-specific antigen* gene (*gag*) encoding a protein shell called a capsid, that encloses the packaged genome. A key difference is that retroviruses also possess an *envelope* gene (*env*), an essential component of infectivity encoding fusogenic glycoproteins that attach and gain entry to target host cells. Yet, *envelope*-like genes appear to have evolved independently multiple times across LTR retrotransposon phylogeny, with the ‘*gypsy viruses*’ of *Drosophila* being one such example. Additionally, many retroviruses fixed in the host germline as endogenous retroviruses have lost their *envelope* gene, making them behave just like LTR retrotransposons. The evolutionary distances spanning LTR retrotransposons and retroviruses are vast, and inferring their phylogenetic relationships is highly challenging. A further layer of complexity arises from the suggestion that LTR retrotransposons may have originated from the fusion of a DNA transposon and a LINE retrotransposon.

Helitron TEs have been suggested to share similarity with prokaryotic rolling-circle replication elements, prokaryotic viruses and plasmids. The monophyly of Crypton TEs and their relationship to DIRS elements (LTR retrotransposons that also utilise tyrosine recombinase to catalyze DNA integration) is unclear. The different superfamilies of DDE

transposons may or may not share a common evolutionary origin, and their history is poorly understood, as is the case for LINE retrotransposons. Things are a little clearer for SINEs, which are believed to have evolved independently on multiple occasions from host transfer RNAs (tRNA), signal recognition particle RNAs (7SL RNA), and 5S rRNAs, after which they acquired additional sequences to promote transposition. However, spare a thought for researchers interested in elucidating the origins of maverick/polinton elements, which are large DNA transposons that may form virions, and which show sequence similarity to DNA viruses, linear plasmids, bacteriophages, adenoviruses, giruses (giant viruses), and transpovirons (plasmid-like elements present in the genomes of giruses)... More generally, the modularity of TE genomes facilitates the potential for shuffling of genes among lineages, and the extent to which such processes have contributed to TE evolution awaits further clarification. Suffice to say, much research remains to elucidate TE origins!

### **Transposable element evolutionary dynamics**

Excluding retroviruses and certain LTR retrotransposons, TEs do not act as infective agents. Instead, their primary transmission route is via host reproduction, as TEs present in host germline cells can transmit vertically to the next generation. Additionally, extensive evidence supports widespread horizontal transmission of transposons (HTT) via passive processes, mediated by parasites and the environment. For example, one such route may be ‘viral hitchhiking’, whereby insertion of a TE into a virus genome facilitates spread to a new host lineage by piggybacking on infective viral transmission. While relatively infrequent, HTT is considered an important mechanism for the survival of many TEs, since persistence in a given host lineage can be short-lived over evolutionary timescales, due to host repression and/or mutational deactivation. Understanding of host range and the determinants of host specificity in TEs is limited in many cases. However, HTT can be viable over large host taxonomic distances, resulting in a web of TE DNA interlinking branches across the tree of life.

In common with many endoparasites, TEs have relatively streamlined genomes, mostly less than 8kb in length. TEs rely on host cellular machinery for their transcription, but they contain genes encoding the proteins necessary for transposition, except for SINEs, which are dependent on enzymes encoded by LINEs. Meanwhile, other TEs can become non-autonomous through mutational inactivation. However, if non-autonomous TEs retain the motifs required for self-recognition by transposition enzymes, they can still transpose by taking advantage of the enzymes expressed by closely related autonomous elements. Where large numbers of non-autonomous TEs build up within a genome, they can outcompete autonomous elements for their own gene products, leading to the eventual demise of the lineage.

Much research remains to elucidate the mechanisms by which TEs mediate their fitness. In certain cases, TEs contain regulatory sequences that promote their expression, or accessory genes involved in host adaptation. Site specificity is poorly described for most TEs, but evidence suggests some show a preference for ‘genomic safe havens’ that reduce the likelihood of incurring negative host impacts. It is also unclear to what extent TE lineages modulate their proliferation rate to switch between high proliferation/high virulence strategies, versus low proliferation/low virulence strategies, that may be less likely to trigger host defences and so lead to improved persistence.

### **The future of transposable element research**

The great accumulation of genomic data over the last two decades has led to a blossoming of TE research. Largescale coordinated genome sequencing initiatives, such as the Earth Biogenome project and the Darwin Tree of Life project, now offer to massively expand both the scale and quality of data available. Extensive sampling of the eukaryotic tree of life substantially increases the power of comparative phylogenetic analyses to disentangle the relative importance of host ecological factors for determining variation in TE distribution. On the other hand, chromosomal-level assemblies permit analyses of within host patterns and processes underlying TE distribution. Meanwhile, a shift from reference genomes towards population-level genomic datasets offers unprecedented opportunities to examine the evolutionary dynamics of TEs, and quantify their contributions to host evolution relative to other forms of genetic variation. Emerging genomic methods also offer to revolutionise our understanding of TE biology at the single-cell level and the finescale mechanisms governing TE proliferation and regulation.

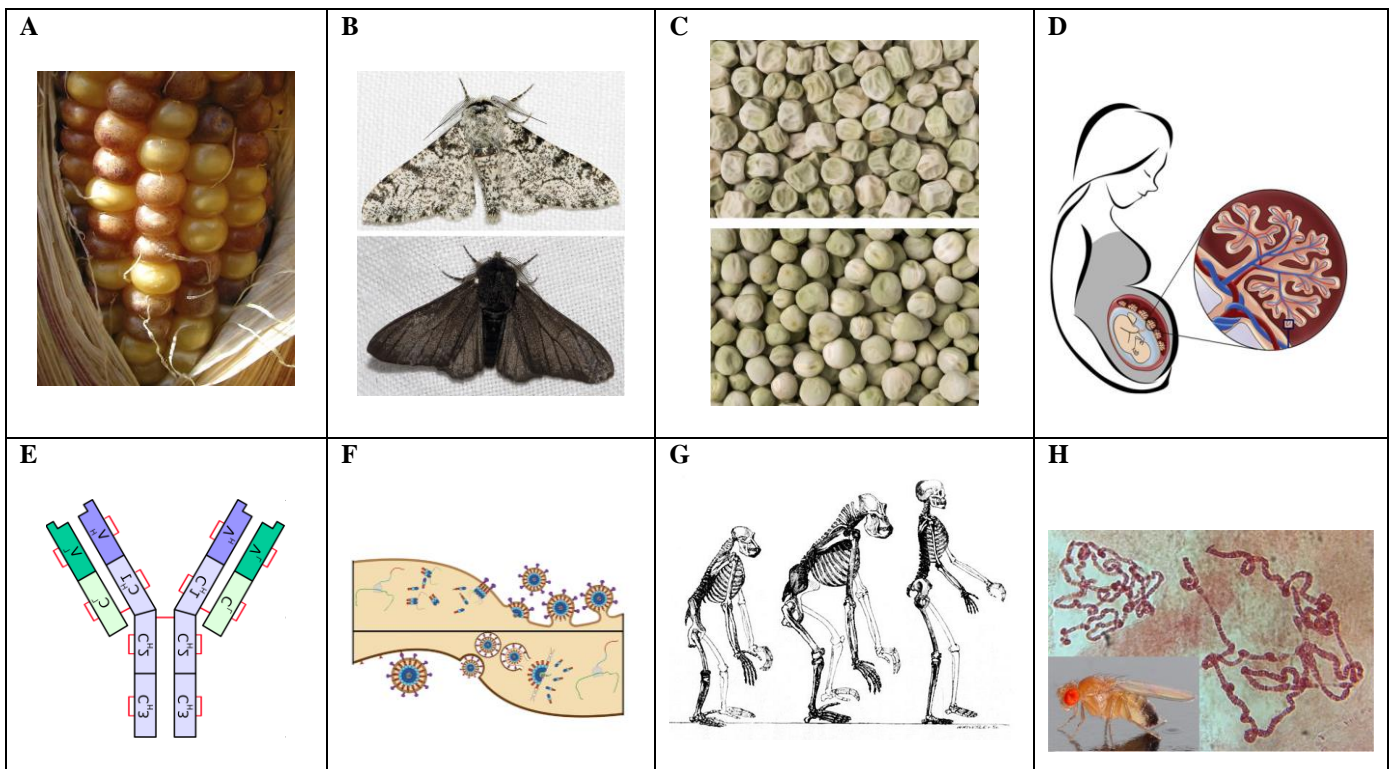
As the quality and quantity of genomic resources increases, it is likely that many further cases of the host utilisation of TE sequences will be uncovered. For example, it was recently demonstrated that a TE provided the genetic mechanism behind the classic case of natural selection in action; industrial melanism in the peppered moth. Accumulating evidence of the involvement of TEs in host evolution raises the question: should TEs be considered fundamental components of host evolvability? Addressing this question holds significance for our understanding of how evolution operates at the genomic level, and by extension, for our comprehension of the mechanisms responsible for generating the earth's rich biodiversity. Progress requires: (i) detailed analyses of the general patterns and processes underlying where, when, and how TEs are recruited for host functions; (ii) quantification of the contributions of TEs relative to other forms of genetic variability. Put another way, if evolution is a tinkerer, does it have a fondness for solving problems with TEs? Affirmation may explain why so few eukaryotic lineages have completely excluded TEs from their genomes, if doing so decreases the ability to respond to selection, thereby increasing extinction risk. For now, it remains an open question to what extent we would see the diversity and complexity of life around us, without the varied contributions of TEs to host evolution.

TEs can be considered the ultimate parasites, operating as they do *within* an organism's genome. Due to their genomic parasitism, we have an unusually rich record of historical data on host-TE dynamics in the form of genomic fossils. We are just beginning to mine this huge record, and there are important issues for the community to address in terms of how data are best shared and stored. However, the research opportunities are vast, and they offer significant potential to shape our understanding of multiple fundamental biological questions, not least the genomic processes underlying eukaryotic evolution.

**Figure 1. Examples of the varied influences of TEs on host traits.**

(A) Pigmentation in maize kernels: A TE insertion disrupts expression of pigmentation, but function is restored in cells where the TE has jumped out (and their daughter cells), resulting in red spotting. Image credit: Demon Lisch, CC BY 2.5, <https://commons.wikimedia.org/w/index.php?curid=1862602>. (B) Melanisation in the peppered moth: TE insertion into an intron of the *cortex* gene increases its expression, leading to wing pattern melanization. Image credit: Chiswick Chap, CC BY-SA 3.0, <https://commons.wikimedia.org/wiki/File:Biston.betularia.f.carbonaria.7209.jpg>, <https://commons.wikimedia.org/w/index.php?curid=880130>. (C) Wrinkled versus smooth peas: Pea wrinkliness was a key trait studied by Mendel in his pioneering work on genetics, and is caused by a TE insertion into the *starch-branching enzyme I* gene. Image credit: The

John Innes Centre. **(D)** Internal gestation in placental mammals: Syncytin genes are co-opted retroviral *envelope* genes that express fusogenic proteins required for the formation of the syncytiotrophoblast layer at the placental materno-fetal interface. Image credit: Lu et al <https://doi.org/10.1016/j.celrep.2017.10.019>. **(E)** Adaptive immunity in jawed vertebrates: RAG proteins originate from co-opted DNA transposons, and play an important role in V(D)J recombination which generates variable antigen binding sites. Image credit: Yohan, CC BY-SA 3.0, <https://commons.wikimedia.org/wiki/File:Anticorps.png>. **(F)** Memory formation: Intercellular signalling for information encoding and storage in the brain involves virus-like properties derived from the *arc* gene, a co-opted retrotransposon *Gag* gene. Image credit: modified from Hantak et al. [https://www.cell.com/trends/neurosciences/fulltext/S0166-2236\(21\)00001-1](https://www.cell.com/trends/neurosciences/fulltext/S0166-2236(21)00001-1). **(G)** Tail-loss in humans and great apes: A SINE inserted into an intron of the *Brachyury* gene pairs with a neighboring SINE in the reverse genomic orientation, which appears to result in an alternative splicing event causing exon-skipping and tail-loss. [Image: Public domain, from Huxley, *Man's Place in Nature*]. **(H)** Telomeres in *Drosophila*: Telomeres are regions of repetitive sequences found at chromosome ends, which protect against DNA damage and fusion with neighbouring chromosomes. *Drosophila* lacks telomerase, the enzyme used by most organisms to maintain telomeres. Instead, active transposition by three classes of non-LTR retrotransposons acts to elongate chromosome ends. Image credit: Chromosome: Nulu iman, CC BY-SA 4.0, [https://commons.wikimedia.org/wiki/File:Chromosome\\_squash\\_from\\_salivary\\_glands\\_of\\_Drosophila\\_melanogaster\\_larvae.png](https://commons.wikimedia.org/wiki/File:Chromosome_squash_from_salivary_glands_of_Drosophila_melanogaster_larvae.png); Fly: André Karwath, CC BY-SA 2.5, [https://en.wikipedia.org/wiki/File:Drosophila\\_melanogaster\\_-\\_side\\_\(aka\).jpg](https://en.wikipedia.org/wiki/File:Drosophila_melanogaster_-_side_(aka).jpg)

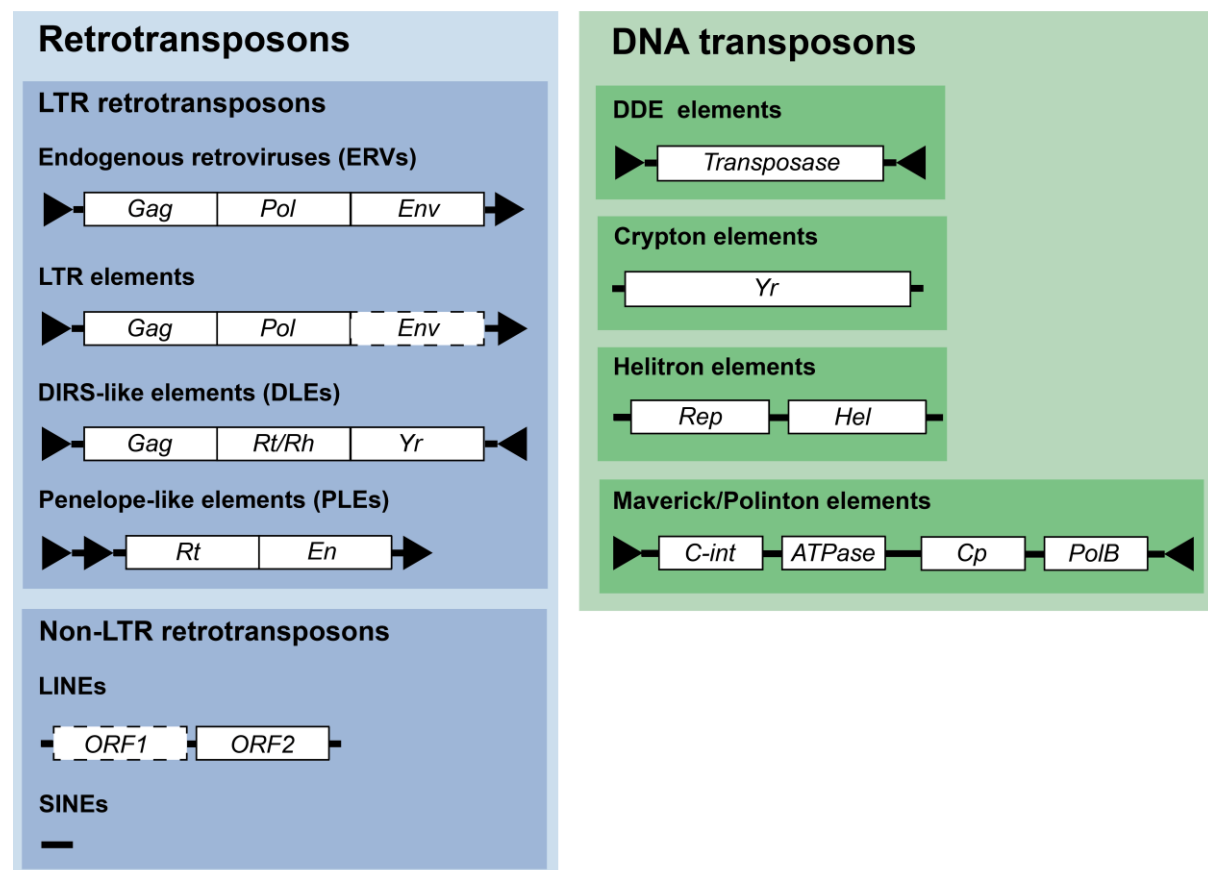


**Figure 2. Generalised schematic of major TE types.**

Retrotransposons are divided into two main groups, long-terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. LTR retrotransposons have direct terminal



repeats, except for DIRS and DIRS-like elements, which have non-identical inverted terminal repeats or split direct terminal repeats respectively. *Gag* encodes core structural proteins that form virus-like particles. *Pol* encodes reverse transcriptase (Rt), a DDE integrase, and a ribonuclease H domain (Rh), which in turn reverse transcribe retrotransposon RNA into DNA, integrate it into the host genome, and cleave phosphodiester bonds between RNA nucleotides in the RNA:DNA hybrid intermediate facilitating synthesis of a double-stranded DNA sequence. *Env* encodes an viral envelope in ERVs, or an envelope-like protein in certain LTR retrotransposons, which enables cell targeting and entry. An endonuclease (En) is used by Penelope-like elements to cleave DNA. A tyrosine recombinase (Yr) is used by DIRS-like elements to perform the role of integration in the place of an integrase (as used in LTR elements and ERVs) or an endonuclease (as used in LINEs and PLEs). Non-LTR retroelements consist of LINEs and SINEs. All LINEs possess an ORF (ORF2) encoding reverse transcriptase (Rt) and Endonuclease (En), and some possess an additional ORF (ORF1) that may encode proteins for functions such as transport of the template RNA to the nucleus. SINEs typically possess a head region derived from an RNA synthesized by RNA polymerase III, a body that often shares similarity to a LINE, and a tail consisting of short simple repeats. DNA transposons are divided into four main types: DDE elements, Crypton elements, Helitron elements and Maverick/Polinton elements. DDE elements and Maverick/Polinton elements both have terminal inverted repeats (TIRs). DDE elements encode a transposase enzyme (Transposase). Crypton elements encode a tyrosine recombinase (Yr) similar to the DIRS-like LTR retrotransposons. Helitron elements encode a HUH endonuclease domain (religation initiator, Rep) and a helicase (Hel) that catalyze rolling circle transposition. Maverick/Polinton elements encode a protein-primed type B DNA polymerase (PolB), a C-intergrase (C-int), a Cysteine protease (Cp), a packaging ATPase, and typically several other genes.



## FURTHER READING

- Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr, Loudet, O., Colot, V., and Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biol.* 22, 138.
- Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 199.
- Burns, K.H. (2020). Our Conflict with Transposable Elements and Its Implications for Human Disease. *Annu. Rev. Pathol.* 15, 51–70.
- Cosby, R.L., Chang, N.-C., and Feschotte, C. (2019). Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* 33, 1098–1116.
- Gilbert, C., and Feschotte, C. (2018). Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr. Opin. Genet. Dev.* 49, 15–24.
- Gilbert, C., Peccoud, J., and Cordaux, R. (2021). Transposable Elements and the Evolution of Insects. *Annu. Rev. Entomol.* 66, 355–372.
- Hayward, A. (2017). Origin of the retroviruses: when, where, and how? *Curr. Opin. Virol.* 25, 23-27.
- Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., et al. (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* 166, 102–114.
- Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., and Heidmann, T. (2013). Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120507.
- Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61.
- Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., et al. (2018). The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* 173, 275.
- Quesneville, H. (2020). Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mobile DNA* 11.
- Ravindran, S. (2012). Barbara McClintock and the discovery of jumping genes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20198–20199.
- Rech, G.E., Radío, S., Guirao-Rico, S., Aguilera, L., Horvath, V., Green, L., Lindstadt, H., Jamilloux, V., Quesneville, H., and González, J. (2022). Population-scale long-read

sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nature Communications* 13.

Robillard, É., Le Rouzic, A., Zhang, Z., Capy, P., and Hua-Van, A. (2016). Experimental evolution reveals hyperparasitic interactions among transposable elements. *Proceedings of the National Academy of Sciences* 113, 14763–14768.

Stitzer, M.C., Anderson, S.N., Springer, N.M., and Ross-Ibarra, J. (2021). The genomic ecosystem of transposable elements in maize. *PLoS Genet.* 17, e1009768.

Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., and Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8, 973–982.

Wells, J.N., and Feschotte, C. (2020). *A Field Guide to Eukaryotic Transposable Elements*. *Annu. Rev. Genet.* 54, 539–561.

Xia, B., Zhang, W., Wudzinska, A., Huang, E., Brosh, R., Pour, M., Miller, A., Dasen, J.S., Maurano, M.T., Kim, S.Y., et al. (2021). The genetic basis of tail-loss evolution in humans and apes. *bioRxiv*, 2021.09.14.460388.